Distal combinatorial tools for graphs of bounded twin-width

Wojciech Przybyszewski ✉ 🏠 📷
Institute of Informatics, University of Warsaw, Poland

──── **Abstract** ────

We study set systems formed by neighborhoods in graphs of bounded twin-width. We start by proving that such graphs have linear neighborhood complexity, in analogy to previous results concerning graphs from classes with bounded expansion and of bounded clique-width. Next, we shift our attention to the notions of distality and abstract cell decomposition, which come from model theory. We give a direct combinatorial proof that the edge relation is distal in classes of ordered graphs of bounded twin-width. This allows us to apply Distal cutting lemma and Distal regularity lemma, so we obtain powerful combinatorial tools for graphs of bounded twin-width.

1 Introduction

Twin-width is a graph width parameter recently introduced by Bonnet et al. in [9] with remarkable properties. In particular, classes of bounded twin-width generalize some of the previously examined graph classes (e.g. planar graphs, graphs excluding a fixed minor, graphs of bounded clique-width), while admitting good structural properties (e.g. being closed under FO-transductions), algorithmic properties (e.g. FO model checking can be done in linear time on graphs of bounded twin-width given a contraction sequence) and combinatorial properties (e.g. χ-boundedness [6]).

Graphs of bounded twin-width have linear neighborhood complexity

Our focus is on the combinatorial and logical complexity of set systems defined by neighborhoods in graphs of bounded twin-width. This continues a line of research studying other graph classes in this context. Namely, graphs of bounded clique-width have linear neighborhood complexity, as it was proven in [26]. That means that for every graph G of clique-width at most c and any non-empty subset A of its vertices there are at most $n_c|A|$ different neighborhoods of vertices of G in A (where the constant n_c depends only on c). The same is also true for planar graphs and graphs excluding a fixed minor. More generally, the same holds for classes with bounded expansion [28], although they are not generalized by

classes of bounded twin-width. Our first result states that graphs of bounded twin-width also have linear neighborhood complexity.

▶ **Theorem** (Informal version of Theorem 4)**.** *For every integer t, there is a constant n_t such that for every graph G of twin-width at most t and every non-empty subset $A \subseteq V(G)$, the vertices of G have at most $n_t|A|$ different neighborhoods in A.*

Our proof relies on properties of matrices of graphs of bounded twin-width shown in [9], as well as the celebrated Stanley-Wilf conjecture/Marcus-Tardos theorem [23]. We remark that Theorem 4 was proven independently by Bonnet et al. in [8]. Although the techniques used there are similar to ours, we decided to include a self-contained proof of Theorem 4 since the underlying ideas are needed in the later part of the paper.

Theorem 4 in particular implies that the VC-density of graphs of bounded twin-width is equal to 1 (Corollary 9). This notion is a refinement of VC-dimension, which is the most well-known measure of complexity of set systems and was introduced in [34]. Both measures are applied in a wide range of fields, including statistical learning theory and computational geometry. Whereas a set system has bounded VC-dimension if and only if it has bounded VC-density, in general, VC-density is at most as large as the VC-dimension, and it turns out that the VC-density, not the VC-dimension, is the decisive measure for the combinatorial complexity of a family of sets [2]. For example, the VC-density of $\mathcal{S}$ governs the size of packings in $\mathcal{S}$ with respect to the Hamming metric [19] and is intimately related to the notions of entropic dimension [3] and discrepancy [25]. VC-density has also a number of applications in algorithmics, for example CONNECTED k-VERTEX COVER admits a kernel with $O(k^{1.5})$ vertices on classes of graphs with VC-density at most 1 [8].

The edge relation is distal in classes of graphs of bounded twin-width

In the following part of the paper we refine our analysis of set systems defined by neighborhoods in graphs of bounded twin-width, and turn our attention to their logical complexity. We do it by investigating the notion of distality defined by Simon in [32]. This notion comes from model theory and aims at describing theories which are NIP and *purely instable*. The original definition is phrased in the language of model theory and was stated for theories and single infinite models. However, there is also an equivalent definition given in [13]. It uses abstract cell decompositions, which are more combinatorial objects. We adapt this definition to the context of classes of finite structures, e.g. graphs., by saying that a formula is distal in a class of structures if it admits an abstract cell decomposition.

It is known that every graph of bounded twin-width can be equipped with a linear order on its vertices such that the twin-width of the ordered graph (seen as a binary structure) is still bounded. However, computing an optimal order is a hard problem. In fact, even the problem of deciding whether a given (unordered) graph has twin-width at most 4 is NP-complete [4]. On the other hand, when it comes to the problem of FO model checking on ordered graphs, it can be solved in fpt time precisely on these classes of ordered graphs that have bounded twin-width [7]. It is also known that classes of ordered graphs of bounded twin-width are distal by combining a few results from model theory. However, this proof is highly unconstructive and relies on the compactness theorem for first-order logic. In the main result of the paper (Theorem 21) we give a direct combinatorial proof that the edge relation is distal in classes of graphs of bounded twin-width, provided that we add to every graph an order that is a witness of a low twin-width.

▶ **Theorem** (Informal version of Theorem 21). *Let $\widehat{\mathcal{C}}$ be a class of ordered graphs of twin-width bounded by some constant t. Then, the formula $\varphi(x;y) \equiv E(x,y)$ admits in $\widehat{\mathcal{C}}$ a distal cell decomposition weakly definable by a finite set $\Psi(x;y_1,\ldots,y_k)$, where $k = \mathcal{O}(t)$. Moreover, the exact form of formulas in Ψ follows from the proof.*

This allows us to e.g. obtain effective bounds on the constants in theorems about powerful distal combinatorial tools proven in [15, 13] and to better understand structure of sets that appear in these theorems.

Distal combinatorial tools

After adapting the notion of abstract cell decompositions to classes of finite structures we translate theorems about powerful combinatorial tools proven in [15, 13] into this setting. We now briefly describe these tools and motivations for investigating abstract cell decompositions.

Cutting lemma

The so-called cutting lemma is a very useful combinatorial partition tool with numerous applications in computational and incidence geometry and related areas (see e.g. [24, Sections 4.5, 6.5] or [11] for a survey). In its simplest form it can be stated as follows (see e.g. [24, Lemma 4.5.3]).

▶ **Theorem** (Cutting lemma). *For every set L of n lines in the real plane and every $1 < r < n$ there exists a $\frac{1}{r}$-cutting for L of size $O(r^2)$. That is, there is a subdivision of the plane into generalized triangles (i.e. intersections of three half-planes) $\Delta_1,\ldots,\Delta_t$ so that the interior of each Δ_i is intersected by at most $\frac{n}{r}$ lines in L, and we have $t \leq Cr^2$ for a certain constant C independent of n and r.*

This result provides a method to analyze intersection patterns in families of lines, and it has many generalizations to higher dimensional sets and/or to families of sets of more complicated shape than lines, for example families of algebraic or semialgebraic curves of bounded complexity [12]. Combining the result of Chernikov, Galvin and Starchenko from [13] with Theorem 21 we get a version of distal cutting lemma for graphs of bounded twin-width.

▶ **Theorem** (Informal version of Theorem 25). *Let $\mathcal{C}$ be a class of graphs of twin-width at most t. Then, there is an integer $d = \mathcal{O}(t)$ with the following property: for any graph $G \in \mathcal{C}$, any $A \subseteq V(G)$ of size n, and any real $1 \leq r \leq n$ we can partition the vertices of $V(G)$ into at most Cr^d sets $X_1,\ldots,X_l$ such that the vertices in every X_i have almost the same neighborhood in A. More precisely, there are at most $\frac{n}{r}$ vertices $a \in A$ for which there are $u,v \in X_i$ with $(u,a) \in E(G)$ and $(v,a) \notin E(G)$.*

Let us remark that this theorem is obvious for $r = 1$ (we just take $X_1 = V(G)$) and the version for $r = n$ corresponds to having a polynomial neighborhood complexity and is implied by Theorem 4.

Regularity Lemma

Another property of distal classes of graphs is a stronger version of the regularity lemma. The original Szemerédi Regularity Lemma is a fundamental result in graph combinatorics with many versions and applications in extremal combinatorics, number theory and computer science (see [20] for a survey). Roughly speaking, it says that for every $\varepsilon > 0$ there is a

bound $K = K(\varepsilon)$ such that every graph G can be partitioned into K parts, such that the edges between any two parts (apart from a few exceptions) behave in a regular way (where ε controls the degree of regularity). In its simplest form it can be presented as follows.

▶ **Theorem** (Regularity lemma). *For every $\varepsilon > 0$ there exists $K = K(\varepsilon)$ such that: for every graph $G = (V, E)$ there exists a partition $V = V_1 \cup \ldots \cup V_k$ into non-empty sets and a set $\Sigma \subseteq [k] \times [k]$ with the following properties.*

1. *Bounded size of the partition: $k \leq K$.*
2. *Few exceptions: $|\bigcup_{(i,j) \in \Sigma} V_i \times V_j| \geq (1 - \varepsilon)|V|^2$.*
3. *ε-regularity: for all $(i, j) \in \Sigma$ and all $A \subseteq V_i, B \subseteq V_j$ with $|A| \geq \varepsilon|V_i|, |B| \geq \varepsilon|V_j|$, one has*

$$|d(A, B) - d(V_i, V_j)| \leq \varepsilon,$$

where $d(X, Y) = \frac{|E(X,Y)|}{|X||Y|}$ and $E(X, Y)$ is the set of edges with one endpoint in X and the other in Y.

In general the bound on the size of the partition K is known to grow as an exponential tower of height $\frac{1}{\varepsilon}$. Recently several improved regularity lemmas were obtained in the context of definable sets in certain structures or in restricted families of structures (see e.g. [33, 22]).

Chernikov and Starchenko proved much stronger version of the regularity lemma for distal structures [15]. For classes of graphs which are NIP (this is a property of classes of structures less restrictive than distality) and their edge relation is distal (in particular graphs of bounded twin-width satisfy this conditions), this can be translated as:

▶ **Theorem** (Informal version of Theorem 20). *Let $\mathcal{C}$ be an NIP class of graphs with a distal edge relation. Then, there is a constant c depending only on $\mathcal{C}$ with the following property: for every $\varepsilon > 0$ and for every graph $G \in \mathcal{C}$, there exists a partition $V(G) = V_1 \cup \ldots \cup V_k$ into non-empty sets, and a set $\Sigma \subseteq [k] \times [k]$ with the following properties.*

1. *Polynomially bounded size of the partition: $k \leq c(\frac{1}{\varepsilon})^c$.*
2. *Few exceptions: $|\bigcup_{(i,j) \in \Sigma} V_i \times V_j| \geq (1 - \varepsilon)|V|^2$.*
3. *$0 - 1$-regularity: for all $(i, j) \in \Sigma$ there are either all edges between V_i and V_j or no edge at all.*

Structure of the paper

The paper is structured as follows. In Section 2 we introduce the preliminary notions. Then, the proof of Theorem 4 is presented in Section 3. After the proof of Theorem 4 we switch our attention to the notion of distality and abstract cell decompositions in Section 4. In particular, we translate the theorems about distal combinatorial tools into the setting of classes of finite structures. Then, in Section 5 we give a direct combinatorial proof that the edge relation is distal in classes of ordered graphs of bounded twin-width. Finally, in Section 6 we present conclusions that follow from our work.

2   Preliminaries

We denote by $[i]$ the set of integers $\{1, \ldots, i\}$. If $\mathcal{X}$ is a set of sets, we denote by $\bigcup \mathcal{X}$ the union of them.

2.1 Graph definitions and notations

In this paper we investigate only undirected simple graphs, i.e. graphs with no multiple-edges nor self-loops. We denote by $V(G)$ the set of vertices of a given graph G and by $E(G)$ the set of its edges.

For a graph $G = (V, E)$ and $A \subseteq V$ we denote by $G - A$ the graph $\{V \setminus A, \{(u, v) \in E : u, v \notin A\}\}$, i.e. the graph obtained from G by removing all the vertices in A and edges incident to them.

For $A \subseteq V(G)$ and $v \in V(G)$ we denote the neighborhood of v in A by $N_A^G(v)$, i.e. $N_A^G(v) = \{u \in A : (u, v) \in E(G)\}$. We omit a superscript G whenever the graph is implicit from the context. We write $N(v)$ for $N_{V(G)}(v)$. Similarly, we denote the set of all neighborhoods in $A \subseteq V(G)$ by $N_G(A)$, i.e. $N_G(A) = \{N_A(u) : u \in V(G)\}$. We also write $\widehat{N_G}(A)$ for the set of all non-empty neighborhoods in A, i.e. $\widehat{N_G}(A) = N_G(A) \setminus \{\emptyset\}$.

For a graph $G = (V, E)$ and two subsets of its vertices $A, B \subseteq V$ we say that A and B are *homogeneous*, if either for every $v \in A$ and $u \in B$ we have $(v, u) \in E$ or for every $v \in A$ and $u \in B$ we have $(v, u) \notin E$. In particular, if $A \cap B \neq \emptyset$ then A and B can be homogeneous only if for every $v \in A$ and $u \in B$ we have $(v, u) \notin E$.

2.2 Matrix definitions and notations

For a matrix M consisting of m rows and n columns (an $m \times n$ matrix) we denote by $M[i][j]$ its entry in the i'th row and j'th column (assuming that $1 \leq i \leq m$ and $1 \leq j \leq n$). A $p \times q$ submatrix N of an $m \times n$ matrix M (with $p \leq m$ and $q \leq n$) is any matrix formed by taking p consecutive rows and q consecutive columns of M. For an $m \times n$ matrix M we denote by $M[i : j][k : l]$ (with $1 \leq i \leq j \leq m$ and $1 \leq k \leq l \leq n$) the submatrix of M formed by rows from i to j (inclusive) and columns from k to l (inclusive).

An $m \times n$ matrix M is *vertical* if every two rows of M are equal, i.e. for any $1 \leq i, j \leq m$ we have $M[i : i][1 : n]$ is equal to $M[j : j][1 : n]$. Similarly, M is *horizontal* if every two columns of M are equal, i.e. for any $1 \leq i, j \leq n$ we have $M[1 : m][i : i]$ is equal to $M[1 : m][j : j]$. Observe that if a matrix is both vertical and horizontal, then it is constant. We say that a matrix is *mixed* if it is neither vertical nor horizontal. A *corner* is a 2×2 mixed matrix. In [9] it was proven that a matrix is mixed if and only if it contains a corner as a submatrix.

Given an $m \times n$ matrix M, a *row-partition* (resp. *column-partition*) is a partition of the rows (resp. columns) of M. Similarly, a *row-division* (resp. *column-division*) is a row-partition (resp. column-partition), where every part consists of consecutive rows (resp. columns). A (k, l)-*division* (or simply division) of a matrix M is a pair $(\mathcal{R}, \mathcal{C})$ of a row-division and a column-division with respectively k and l parts. A *zone* of a division $(\mathcal{R}, \mathcal{C}) = (\{R_1, \ldots, R_k\}, \{C_1, \ldots, C_l\})$ is any submatrix $R_i \cap C_j$ for $1 \leq i \leq k$ and $1 \leq j \leq l$.

A $0, 1$-matrix is a matrix with all its entries equal to 0 or 1. Given a $0, 1$-matrix M, a t-*grid minor* in M is a (t, t)-division of M in which every zone contains a 1. Given a $0, 1$-matrix M, a t-*mixed minor* in M is a (t, t)-division of M in which every zone is a mixed submatrix of M. A matrix is t-*grid free* (resp. t-*mixed free*) if it does not contain a t-grid minor (resp. t-mixed minor).

If G is an n-vertex graph and σ is a total ordering of $V(G)$, say, $v_1, \ldots, v_n$, then $M_\sigma(G)$ denotes the adjacency matrix of G in the order σ. Thus $M_\sigma(G)[i][j]$ is 1 if $(v_i, v_j) \in E(G)$ and 0 otherwise. An *ordered graph* is a pair $(G, \preceq)$ such that G is a graph and $\preceq$ is a total order on $V(G)$. We say that an ordered graph $(G, \preceq)$ is t-mixed free if $M_\preceq(G)$ is t-mixed free. A first-order formula of ordered graphs is a standard first-order formula of graphs that

can use additional $\preceq$ symbol, which is interpreted as an order of an ordered graph.

2.3   Definition of twin-width of graphs

The twin-width of a given graph was first defined in [9]. Since then, many tutorials on the topic were given, e.g. [17, 27] can serve as a good introduction.

The original definition of twin-width uses the notion of a *trigraph*, i.e. a triple $G = (V, E, R)$ where E and R are two disjoint sets of edges on V: the (usual) edges and the red edges. A trigraph (V, E, R) such that (V, R) has maximum degree at most d is a d-trigraph. Any graph (V, E) may be interpreted as the trigraph $(V, E, \emptyset)$.

Given a trigraph $G = (V, E, R)$ and two vertices u, v in V, we define the trigraph $G/u, v = (V', E', R')$ obtained by identifying u, v into a new vertex w as the trigraph on a new vertex-set $V' = (V \setminus \{u, v\}) \cup \{w\}$ such that $G - \{u, v\} = (G/u, v) - \{w\}$ and the following edges incident to w:

- $(w, x) \in E'$ if and only if $(u, x) \in E$ and $(v, x) \in E$,
- $(w, x) \notin E' \cup R'$ if and only if $(u, x) \notin E$ and $(v, x) \notin E$,
- $(w, x) \in R'$ otherwise.

We say that $G/u, v$ is a *contraction* of G. If both G and $G/u, v$ are d-trigraphs, $G/u, v$ is a d-contraction.

A (tri)graph G is d-collapsible if there exists a sequence of d-contractions which contracts G to a single vertex. The minimum d for which G is d-collapsible is the *twin-width* of G, denoted $\mathrm{tww}(G)$.

2.4   Theorems related to twin-width of graphs

The following grid minor theorem for twin-width proven in [9] describes a connection between twin-width of a graph and mixed minors of its matrix.

▶ **Theorem 1.** *If G is a graph of twin-width less than t, then there is a total ordering on its vertices σ such that $M_\sigma(G)$ is $2t + 2$-mixed free. On the other hand, if G is a graph and σ is a total ordering on its vertices such that $M_\sigma(G)$ is k-mixed free, then $\mathrm{tww}(G) = 2^{2^{O(k)}}$.*

A crucial tool for this proof is the Marcus-Tardos theorem about grid minors.

▶ **Theorem 2** ([23]). *For every integer t, there is some c_t such that every $m \times n$ $0, 1$-matrix M with at least $c_t \max(m, n)$ entries 1 has a t-grid minor.*

Marcus and Tardos established Theorem 2 with $c_t = 2t^4\binom{t^2}{t}$. Then Cibulka and Kynčl [16] decreased c_t to $8/3(t + 1)^2 2^{4t}$. In the rest of this paper we keep the notation c_t for these constants.

2.5   Vapnik-Chervonenkis density of graphs

In this section we introduce notions of Vapnik-Chervonenkis dimension (VC-dimension) and Vapnik-Chervonenkis density (VC-density) in the context of graphs. It is worth noting that it was first introduced in [34] and originated in statistics, but it is also an important notion in combinatorics and statistical learning theory.

We say that $A \subseteq V(G)$ is *shattered* if $N_G(A) = 2^A$, i.e. for every subset of A there is a vertex $v \in V(G)$ such that $N_A(v)$ is precisely that subset of A. For a graph G we define its
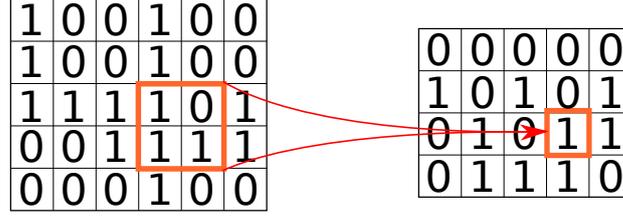
Figure 1 Example of a corner matrix

VC-dimension as the size of the largest subset of $V(G)$, which is shattered, and denote it by $\dim(G)$. We denote the *shatter function* of a graph G by $\pi_G : \mathbb{N} \to \mathbb{N}$ and define it as follows:

$$\pi_G(k) = \max_{A \subseteq V(G), |A| \leq k} |N_G(A)|.$$

For a non-empty class of graphs $\mathcal{C}$ we define its VC-dimension as the supremum over VC-dimensions of graphs in $\mathcal{C}$ and denote it by $\dim(\mathcal{C})$. Similarly, we denote its shatter function by $\pi_{\mathcal{C}} : \mathbb{N} \to \mathbb{N}$ and define it by:

$$\pi_{\mathcal{C}}(k) = \max\{\pi_G(k) : G \in \mathcal{C}\}.$$

Observe that this is well-defined, as $\pi_G(k) \leq 2^k$ for any graph G and $k \in \mathbb{N}$.

We can bound the shatter function of a graph class $\mathcal{C}$ in terms of $\dim(\mathcal{C})$. This is stated in the shatter function lemma proven independently by Sauer [29] and Shelah [30].

▶ **Lemma 3** (Shatter function lemma). *If $\mathcal{C}$ is a graph class such that $\dim(\mathcal{C}) \leq d$, then $\pi_{\mathcal{C}}(k) = O(k^d)$.*

For a class of graphs $\mathcal{C}$ we also define its *VC-density* (denoted by $\mathrm{vc}(\mathcal{C})$) as follows:

$$\mathrm{vc}(\mathcal{C}) = \begin{cases} \inf\{r \in \mathbb{R} : r > 0, \pi_{\mathcal{C}}(k) = O(k^r)\} & \text{if } \dim(\mathcal{C}) \text{ is finite,} \\ +\infty & \text{otherwise.} \end{cases}$$

3 Neighborhood complexity of classes of graphs of bounded twin-width

In this section we prove that classes of graphs of bounded twin-width admit linear neighborhood complexity. This is formalized in Theorem 4.

▶ **Theorem 4.** *For every integer t, there is some n_t such that for every graph G of twin-width at most t and every non-empty $A \subseteq V(G)$ we have $|N_G(A)| \leq n_t|A|$.*

The following definition and two lemmas will be useful for the proof.

▶ **Definition 5.** *Let A be an $m \times n$ $0,1$-matrix, where $m, n \geq 2$. We define **the corner matrix** B **of matrix** A as the $(m-1) \times (n-1)$ matrix given by:*

$$B[i][j] = \begin{cases} 1 & \text{if } M[i:(i+1)][j:(j+1)] \text{ is a corner,} \\ 0 & \text{otherwise.} \end{cases}$$

An example of a corner matrix is presented in Figure 1.
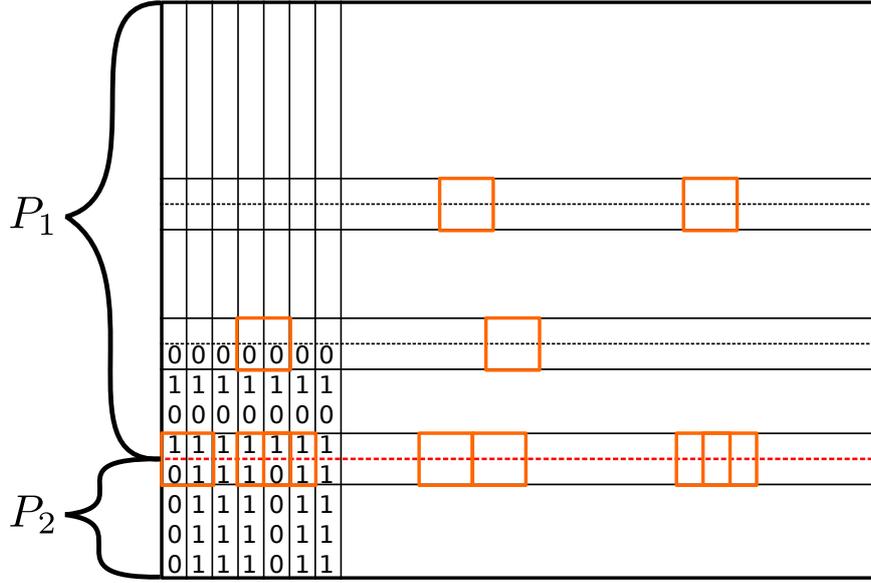
Figure 2 A matrix P with all its corners (in orange) contained in three pairs of consecutive rows is split into P_1 and P_2 in the inductive step.

▶ **Lemma 6.** *Let A be a $0, 1$-matrix, which is t-mixed free. Then its corner matrix B is $2t$-grid free.*

Proof. Let us assume by contradiction, that B admits a $(2t, 2t)$-division $(\mathcal{R}, \mathcal{C}) = (\{R_1, \ldots, R_{2t}\}, \{C_1, \ldots, C_{2t}\})$, which is a $2t$-grid minor. Every part R_k, which consists of l rows of B, corresponds to $l + 1$ consecutive rows of A. Precisely, if R_k consists of rows i to $i + l - 1$ of B, then entries in these rows depend only on rows i to $i + l$ of A. Moreover, whenever $|k - m| > 1$, R_k and R_m correspond to disjoint sets of rows of A – in particular $R_1, R_3, \ldots, R_{2t-1}$ correspond to t disjoint sets of rows of A. The same happens for $C_1, C_3, \ldots, C_{2t-1}$. These sets of rows and columns induce a t-mixed minor of A. This is a contradiction to A being t-mixed free. ◀

▶ **Lemma 7.** *Let P be an $m \times n$ $0, 1$-matrix such that all corners in P appear only in p different pairs of consecutive rows, i.e. there are p distinct integers $1 \leq i_1, \ldots, i_p < m$ such that every corner in P is of a form $P[i_j : (i_j + 1)][k : (k + 1)]$ for some $1 \leq j \leq p$ and $1 \leq k < n$. Then P has at most 2^{p+1} different columns.*

Proof. We proceed by induction on p. In the base case, i.e. when $p = 0$, P does not contain any corner, so it is either horizontal or vertical. If P is horizontal, then all its columns are equal by definition. On the other hand, when P is vertical, then every column of P contains only 0s or only 1s. Therefore, P contains at most 2 different columns.

Now let us assume $p \geq 1$ and the statement is true for $p - 1$. Without loss of generality $1 \leq i_1 < \ldots < i_p < m$. We can split P horizontally into two submatrices $P_1 = P[1 : i_p][1 : n]$ and $P_2 = P[(i_p + 1) : m][1 : n]$. All the corners of P_1 are contained in $p - 1$ different pairs of consecutive rows, so it contains at most 2^p different columns (see Figure 2 for an illustration). Moreover, P_2 does not contain any corner at all, so it contains at most 2 different columns. Every column of P is made up using one column of P_1 and one column of P_2, so there are at most $2^p \cdot 2 = 2^{p+1}$ different columns in P. ◀

We also state an easy corollary which follows directly from the proof of Lemma 7 and will be useful in later sections.

▶ **Corollary 8.** *Let P be an $m \times n$ $0, 1$-matrix such that all corners in P appear only in p different pairs of consecutive rows. Then, there is a subset $I \subseteq [m]$ of at most $p + 1$ rows such that whenever two columns j and j' satisfy $P[i][j] = P[i][j']$ for every $i \in I$, then they are equal.*

Now we prove that graphs of bounded twin-width have linear neighborhood complexity.

Proof of Theorem 4. Let us take any graph G such that $\text{tww}(G) \leq t$ and a non-empty subset $A \subseteq V(G)$ of vertices of G. By Theorem 1 there is a total ordering σ on $V(G)$ such that $M_\sigma(G)$ is $2t + 2$-mixed free. For the rest of this proof we write M for $M_\sigma(G)$.

Let $B \subseteq V(G)$ be a minimal subset of vertices of G such that they represent all possible neighborhoods in A apart from $\emptyset$. Formally, B is a minimal subset of $V(G)$ such that $\widehat{N_G(A)} = \{N_A(b) : b \in B\}$. Obviously $|N_G(A)| \leq |B| + 1$ so we need to bound the size of $|B|$. Let us also remark, that for any distinct $b, b' \in B$ we have $N_A(b) \neq N_A(b')$, as in the opposite case we could remove b' from B obtaining a smaller set with the property we want. Similarly, for any $b \in B$ we have $N_A(b) \neq \emptyset$.

Let us denote by N the matrix that is obtained from M by removing rows that correspond to vertices in $V(G) \setminus A$ and columns that correspond to vertices in $V(G) \setminus B$. Obviously, N is an $|A| \times |B|$ $2t + 2$-mixed free matrix, as we obtain it by deleting some rows and columns of M, which is $2t + 2$-mixed free itself. We can see that any two consecutive columns of N form a mixed matrix. Indeed, they cannot be equal, so the matrix formed by them is not vertical. It is not horizontal either, as there is no column in N containing only zeros. Therefore, for any two consecutive columns of N there is a corner contained in them. Let us denote by C the corner matrix of N (we assume $|A|, |B| \geq 2$ as the opposite case is trivial). From what we observed so far, we know that in every column of C there is at least one entry with 1.

By Lemma 7 applied for $p = c_{4t+4}$ we obtain that every set of $2^{c_{4t+4}}$ consecutive columns of C has at least c_{4t+4} non-zero rows. This is because every such set corresponds to a set of $2^{c_{4t+4}} + 1$ consecutive columns in N (and columns of N are pairwise different).

Suppose C has at least $2^{c_{4t+4}}(|A| - 1)$ columns. Pick any partition of its columns into $|A| - 1$ disjoint sets $(B_1, \ldots, B_{|A|-1})$, of at least $2^{c_{4t+4}}$ consecutive columns each. Consider the $(|A| - 1) \times (|A| - 1)$ matrix C' given by:

$$C'[i][j] = \begin{cases} 1 & \text{if there is a column in } B_j \text{ that has entry 1 in } i\text{'th row,} \\ 0 & \text{otherwise.} \end{cases}$$

The matrix C' has at least c_{4t+4} entries with 1 in every column. Therefore, by Theorem 2, it admits a $4t + 4$-grid minor, which trivially induces a $4t + 4$-grid minor of C. This is a contradiction with Lemma 6, which states that C is $4t + 4$-grid free. Therefore C has at most $2^{c_{4t+4}}(|A| - 1) - 1$ columns, so G has at most $2^{c_{4t+4}}(|A| - 1) + 1$ neighborhoods in $|A|$. Bearing in mind the separate case for $|A| = 1$, we can set $n_t = 2^{c_{4t+4}} = 2^{2^{O(t)}}$. ◀

▶ **Corollary 9.** *Let $\mathcal{C}_t$ be a class of all graphs of twin-width at most t. Then $\text{vc}(\mathcal{C}_t) = 1$.*

Proof. Theorem 4 trivially implies $\text{vc}(\mathcal{C}_t) \leq 1$, so it suffices to show $\text{vc}(\mathcal{C}_t) \geq 1$. By G_k we denote the graph on vertices $[2k]$ with edges $\{(2i - 1, 2i) : i \in [k]\}$, i.e. G_k is a sum of k non-incident edges. Let us also take $A_k = \{2i : i \in [k]\}$. Obviously G_k has twin-width 0, so $G_k \in \mathcal{C}_t$ for any $t \geq 0$. On the other hand, $|N_{G_k}(A_k)| = |A_k| + 1$, which implies $\text{vc}(\mathcal{C}_t) \geq 1$. ◀

4   Distality and distal combinatorial tools

In this section we concentrate on the notion of *distality*. The definition of a distal theory was introduced in [32] using notions from model theory. Later in [13], an equivalent definition of a distal model was presented, this time using *abstract cell decompositions*, which are more combinatorial objects. To our knowledge, only the definitions of a distal theory and a distal model are explicitly present in the literature. In this section, we also define notions of a distal class of structures and a distal formula. We remark, that they are just straightforward generalizations of the previous definitions and they already appeared in the literature implicitly (e.g. [13, Theorem 3.2]). Using our definitions we reformulate theorems about powerful combinatorial tools presented in [13] and [15]. Originally, they were stated for single, infinite distal structures, but we derive analogous statements for the case of classes of (possibly finite) structures. However, we need to start with additional preliminaries on model theory and introduce the notion of *NIP* (standing for *not the independence property*), which is closely related to distality.

4.1   Model theory preliminaries

In the remaining part of this paper we investigate only first-order structures over relational signatures. In particular we treat a graph G as a structure over the signature consisting only of the binary edge relation symbol E. Similarly, we treat an ordered graph $(G, \preceq)$ as a structure over the signature consisting of two binary symbols – the edge relation symbol E and the order relation symbol $\preceq$.

Our notations are standard. We usually denote first-order structures by blackboard bold letters $\mathbb{M}, \mathbb{N}$, etc., and use letters M, N, etc. to denote their domains. We denote elements of domains and variables in formulas by small letters a, x, etc., and use small letters with a bar $\overline{a}, \overline{x}$, etc. to denote tuples of elements or variables. Then we use $|\overline{x}|$ to denote the length of the tuple $\overline{x} = (x_1, \ldots, x_n)$.

We sometimes treat a tuple $\overline{x}$ as a finite set of variables, and then we write $\varphi(\overline{x})$ to denote a first-order formula φ with free variables contained in $\overline{x}$. We may also write $\varphi(\overline{x}_1, \ldots, \overline{x}_k)$ to denote a formula whose free variables are contained in $\overline{x}_1 \cup \ldots \cup \overline{x}_k$. We write x instead of $\{x\}$ in case of a singleton set of variables, e.g. $\varphi(x, y)$ always refers to a formula with two free variables x and y. We sometimes write $\varphi(\overline{x}; \overline{y})$ (with a semicolon between tuples) to distinguish a partition of the set of free variables of φ into two parts, $\overline{x}$ and $\overline{y}$; this partition plays an implicit role in some definitions. If we have a set Ψ of formulas with the same free variables $\overline{x}$ (i.e. $\Psi = \{\varphi_1(\overline{x}), \varphi_2(\overline{x}), \ldots\}$), then we may denote this set by $\Psi(\overline{x})$ to emphasize the set of free variables.

If $\mathbb{M}$ is a structure and $\varphi(\overline{x}; \overline{y})$ is a formula in the language of $\mathbb{M}$, then for $\overline{a} \in M^{|\overline{y}|}$, by $\varphi(M; \overline{a})$ we denote the subset of $M^{|\overline{x}|}$ defined by $\varphi(\overline{x}; \overline{a})$, namely $\varphi(M; \overline{a}) = \{\overline{b} \in M^{|\overline{x}|} : \mathbb{M} \models \varphi(\overline{b}; \overline{a})\}$.

For a given signature Σ we say that a Σ-structure $\mathbb{A}$ is a *binary structure*, if Σ is a signature consisting only of binary relational symbols.

A *theory* T (over Σ) is a set of Σ-sentences. A *model of* a theory T is a Σ-structure $\mathbb{M}$ such that $\mathbb{M} \models \varphi$ for all $\varphi \in T$. When a theory has a model, it is said to be *consistent*. The *theory of* a class of Σ-structures $\mathcal{C}$ is the set of all Σ-sentences φ such that $\mathbb{M} \models \varphi$ for all $\mathbb{M} \in \mathcal{C}$. The *elementary closure* $\overline{\mathcal{C}}$ of $\mathcal{C}$ is the set of all models $\mathbb{M}$ of the theory of $\mathcal{C}$. Thus $\mathcal{C} \subset \overline{\mathcal{C}}$, and $\mathcal{C}$ and $\overline{\mathcal{C}}$ have equal theories.

An important tool for checking if a theory is consistent is the compactness theorem.

▶ **Theorem 10** (Compactness). *A theory T is consistent if and only if every finite subset T' of T is consistent.*

4.2 NIP and monadic NIP

The notion of *NIP* formulas was introduced by Shelah in his work on the classification program [31]. We present here a definition from [21], which is equivalent to the original one.

▶ **Definition 11.** *A formula $\varphi(\overline{x}; \overline{y})$ with two tuples of free variables $\overline{x}$ and $\overline{y}$ is NIP (or dependent) in a class of structures $\mathcal{C}$ if there is a constant k, such that for every structure $\mathbb{M} \in \mathcal{C}$ there are no tuples $\overline{a_i} \in M^{|\overline{x}|} (i \in [k])$, $\overline{b_I} \in M^{|\overline{y}|} (I \subseteq [k])$ satisfying*

$$\mathbb{M} \models \varphi(\overline{a_i}; \overline{b_I}) \iff i \in I. \tag{$*$}$$

Moreover, we say that a class of structures $\mathcal{C}$ is NIP if every formula $\varphi(\overline{x}; \overline{y})$ is NIP in $\mathcal{C}$ and we say that a single structure $\mathbb{M}$ is NIP if the class $\{\mathbb{M}\}$ is.

Observe, that for a formula $\varphi(\overline{x}; \overline{y})$ and a constant k we can write a first-order sentence $\mathrm{assert}(\varphi(\overline{x}; \overline{y}), k)$ which states, that in a given structure $\mathbb{M}$ there are no tuples $\overline{a_i} \in M^{|\overline{x}|} (i \in [k])$, $\overline{b_I} \in M^{|\overline{y}|} (I \subseteq [k])$ satisfying condition $(*)$, namely

$$\mathrm{assert}(\varphi(\overline{x}; \overline{y}), k) \equiv \neg \left[\exists_{\overline{x}_1} \dots \exists_{\overline{x}_k} \exists_{\overline{y}_\emptyset} \dots \exists_{\overline{y}_{[k]}} \left(\bigwedge_{i \in I} \varphi(\overline{x}_i; \overline{y}_I) \wedge \bigwedge_{i \notin I} \neg \varphi(\overline{x}_i; \overline{y}_I) \right) \right].$$

Therefore, if $\varphi(\overline{x}; \overline{y})$ is NIP in a class of structures $\mathcal{C}$, then for some k the formula $\mathrm{assert}(\varphi(\overline{x}; \overline{y}), k)$ is true in all structures in $\mathcal{C}$. It follows that $\varphi(\overline{x}; \overline{y})$ is NIP in $\mathcal{C}$ if and only if it is NIP in its elementary closure $\overline{\mathcal{C}}$. Moreover, $\mathcal{C}$ is an NIP class of structures if and only if $\overline{\mathcal{C}}$ is.

We remark that the definition of NIP formulas can be equivalently expressed in terms of VC-dimension. Indeed, it is easy to see that a formula $\varphi(\overline{x}; \overline{y})$ is NIP in a class of structures $\mathcal{C}$ if and only if there is an integer k such that for every $\mathbb{M} \in \mathcal{C}$ the family of sets $\{\varphi(M; \overline{a}) : \overline{a} \in M^{|y|}\}$ has VC-dimension at most k. This gives an equivalent definition of NIP using the notion of VC-dimension.

Let us also note at this point that a class of graphs of bounded twin-width together with compatible orders is NIP. We state it formally in the following lemma.

▶ **Lemma 12.** *Let $\widehat{\mathcal{C}}$ be a class of ordered graphs with the following property: there is a constant $t \in \mathbb{N}$ depending only on $\widehat{\mathcal{C}}$ such that for every $(G, \preceq) \in \widehat{\mathcal{C}}$ the matrix $M_\preceq(G)$ is t-mixed free. Then, $\widehat{\mathcal{C}}$ is NIP.*

We defer the proof of Lemma 12 to Appendix A. The reason is that although the proof is a simple combination of [9, Theorem 14], [7, Theorem 11], and [7, Theorem 3], it requires introducing a number of new definitions which are irrelevant for the rest of this paper.

4.3 Abstract cell decompositions and definitions of distality

In this subsection we define the notion of an *abstract cell decomposition* for a given formula in a given structure by following [13]. This requires a number of new definitions – in order to make it easier to understand them, we interleave new definitions with examples. Finally, we use abstract cell decompositions for defining the notion of distality.

Let $\mathbb{M}$ be a fixed Σ-structure. For sets $A, X \subseteq M^d$ we say that *A crosses X* if both $X \cap A$ and $X \cap (M^d \setminus A)$ are nonempty.

For a formula $\varphi(\overline{x}; \overline{y})$ and a set $S \subseteq M^{|\overline{y}|}$ we say that a subset $A \subseteq M^{|\overline{x}|}$ is *$\varphi(\overline{x}; S)$-definable* if $A = \varphi(M; \overline{s})$ for some $\overline{s} \in S$. For a set $A \in M^{|\overline{x}|}$ we say that *$\varphi(\overline{x}; S)$ crosses A* if $\varphi(M; \overline{s})$ crosses A for some $\overline{s} \in S$.
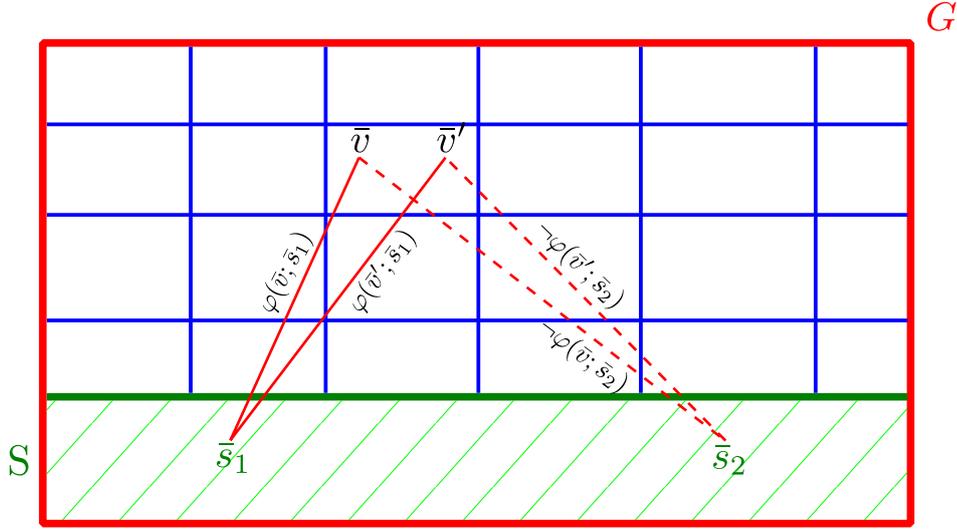
■ **Figure 3** For a given finite set $S \subseteq M^{|y|}$ we can divide $M^{|x|}$ into at most $2^{|S|}$ parts such that tuples in every part satisfy φ with the same elements of S. That gives us an abstract cell decomposition for $\varphi(\overline{x}; S)$. However, this decomposition does not need to be weakly definable.

▶ **Example 13.** Consider the structure $\mathbb{M}$ with domain $M = \{v_1, v_2, \ldots\}$ over the empty signature $\Sigma = \emptyset$ and a formula $\varphi(x; y) \equiv x = y$. Then, for $S_k = \{v_1, \ldots, v_k\}$ only $\{v_1\}, \ldots, \{v_k\}$ are $\varphi(x; S)$-definable. Moreover, $\varphi(x; S_k)$ crosses $\{v_1, v_2\}$ (because $\{v_1\}$ does), but it does not cross $M \setminus S_k$.   ◀

Given a finite set $S \subseteq M^{|\overline{y}|}$, a finite family $\mathcal{F}$ of subsets of $M^{|\overline{x}|}$ is called an *abstract cell decomposition* for $\varphi(\overline{x}, S)$ if $M^{|\overline{x}|} = \bigcup \mathcal{F}$ and every $\Delta \in \mathcal{F}$ is not crossed by $\varphi(\overline{x}; S)$. Observe, that we do not require the sets in $\mathcal{F}$ to be disjoint – we just want them to cover the whole of $M^{|\overline{x}|}$. An abstract cell decomposition for $\varphi(\overline{x}; \overline{y})$ is an assignment $\mathcal{T}$ that to each finite set $S \subseteq M^{|\overline{y}|}$ assigns an abstract cell decomposition $\mathcal{T}(S)$ for $\varphi(\overline{x}; S)$.

Observe that every $\varphi(\overline{x}; \overline{y})$ admits an obvious abstract cell decomposition, with $\mathcal{T}(S)$ consisting of the atoms in the Boolean algebra generated by the $\varphi(\overline{x}; S)$-definable sets (see Figure 3). In general, defining these cells would require longer and longer formulas when S grows, and we want to avoid this possibility. Therefore, we say that an abstract cell decomposition $\mathcal{T}$ for $\varphi(\overline{x}; \overline{y})$ is *weakly definable* if there is a finite set of formulas $\Psi(\overline{x}; \overline{y_1}, \ldots, \overline{y_k})$ with $|\overline{y}| = |\overline{y_1}| = \ldots = |\overline{y_k}|$ such that for any finite $S \subseteq M^{|\overline{y}|}$, every $\Delta \in \mathcal{T}(S)$ is $\Psi(\overline{x}; S^k)$-definable (i.e. $\Delta = \psi(M; \overline{s_1}, \ldots, \overline{s_k})$ for some $\overline{s_1}, \ldots, \overline{s_k} \in S$ and $\psi \in \Psi$). In this case we say that $\Psi(\overline{x}; \overline{y_1}, \ldots, \overline{y_k})$ weakly defines $\mathcal{T}$.

▶ **Example 14.** Consider the structure $\mathbb{M}$ and the formula $\varphi(x; y)$ from Example 13. The obvious abstract cell decomposition for $\varphi(x; y)$ mentioned above is the assignment $\mathcal{T}$ defined for every finite $S \subseteq M$ as follows

$$\mathcal{T}(S) = \{\{v\} : v \in S\} \cup \{M \setminus S\}.$$

However, $\varphi(x; y)$ does not admit a weakly definable abstract cell decomposition in $\mathbb{M}$. Indeed, assume by contradiction that $\varphi(x; y)$ admits an abstract cell decomposition $\mathcal{T}$ in $\mathbb{M}$ weakly definable by $\Psi(x; y_1, \ldots, y_k)$. Take a set S of size $k + 1$ and an element $v \in M \setminus S$. By the definition of $\mathcal{T}$ there is a set $F \in \mathcal{T}(S)$ such that $v \in F$. This set F is definable by some $\psi \in \Psi$, so $F = \psi(M; s_1, \ldots, s_k)$ for some $s_1, \ldots, s_k \in S$. Therefore, there is an element

$s \in S \setminus \{s_1, \ldots, s_k\}$. As ψ is a formula over the empty signature, it cannot distinguish v from s, so $s \in F$. However, $\{s\}$ crosses F, which is a contradiction. ◄

▶ **Example 15.** We observed in Example 14 that the formula which is just an equality does not admit a weakly definable abstract cell decomposition in an infinite structure over the empty signature. We can overcome this problem by adding a total order to our structure. Formally, consider the structure $\mathbb{M}_{\preceq}$ over the signature $\Sigma = \{\preceq\}$ with the domain $M = \{v_1, v_2, \ldots\}$ and assume $v_i \preceq v_j \iff i \leq j$. Consider an abstract cell decomposition $\mathcal{T}$ defined for any $S = \{v_{i_1}, \ldots, v_{i_k}\}$ with $i_1 < \ldots < i_k$ as follows

$$\mathcal{T}(S) = \{\{v\} : v \in S\}$$
$$\cup \{\{v : v_{i_j} \prec v \prec v_{i_{j+1}}\} : j \in [k-1]\}$$
$$\cup \{\{v : v \prec v_{i_1}\}\}$$
$$\cup \{\{v : v \succ v_{i_k}\}\}.$$

Observe that every $F \in \mathcal{T}(S)$ can be defined by one of these four formulas:

- $\psi_1(x; y_1) \equiv x = y_1$;
- $\psi_2(x; y_1, y_2) \equiv y_1 \preceq x \wedge x \preceq y_2 \wedge x \neq y_1 \wedge x \neq y_2$;
- $\psi_3(x, y_1) \equiv x \preceq y_1 \wedge x \neq y_1$;
- $\psi_4(x; y_1) \equiv y_1 \preceq x \wedge x \neq y_1$.

Formally, these formulas have a different number of parameters, but we can extend ψ_1, ψ_3 and ψ_4 with one artificial parameter which is not further used and in this way we obtain that $\mathcal{T}$ is weakly definable with $\Psi(x; y_1, y_2)$ for $\Psi = \{\psi_1, \psi_2, \psi_3, \psi_4\}$. ◄

If we have an abstract cell decomposition $\mathcal{T}$ which is weakly definable by $\Psi(\overline{x}, \overline{y_1}, \ldots, \overline{y_k})$ we know that for every $F \in \mathcal{T}(S)$ there is a choice of a formula $\psi \in \Psi$ and $\overline{s_1}, \ldots, \overline{s_k} \in S$ such that $F = \psi(\mathbb{M}, \overline{s_1}, \ldots, \overline{s_k})$. However, we don't know *a priori* which choices of $\psi \in \Psi$ and parameters from S lead to a set in $\mathcal{T}(S)$. Therefore, we say that an abstract cell decomposition $\mathcal{T}$ for $\varphi(\overline{x}; \overline{y})$ is *definable* if it is weakly definable by some $\Psi(\overline{x}; \overline{y_1}, \ldots, \overline{y_k})$ and if for every $S \subset M^{|\overline{y}|}$ and each $\Psi(\overline{x}; S^k)$-definable $\Delta \subseteq M^{|\overline{x}|}$ there is a set $\mathcal{I}(\Delta) \subseteq M^{|\overline{y}|}$, *uniformly definable* in Δ, such that

$$\mathcal{T}(S) = \{\Delta \in \Psi(\overline{x}; S) : \mathcal{I}(\Delta) \cap S = \emptyset\}.$$

By the uniform definability of $\mathcal{I}(\Delta)$ we mean that for every $\psi(\overline{x}; \overline{y_1}, \ldots, \overline{y_k}) \in \Psi(\overline{x}; \overline{y_1}, \ldots, \overline{y_k})$, there is a formula $\theta_\psi(\overline{y}; \overline{y_1}, \ldots, \overline{y_k})$ such that for any $\overline{s_1}, \ldots, \overline{s_k} \in M^{|\overline{y}|}$ if $\Delta = \psi(M; \overline{s_1}, \ldots, \overline{s_k})$ then $\mathcal{I}(\Delta) = \theta_\psi(M; \overline{s_1}, \ldots, \overline{s_k})$.

Observe, that every abstract cell decomposition $\mathcal{T}$ weakly definable with $\Psi(\overline{x}; \overline{y_1}, \ldots, \overline{y_k})$ induces the definable abstract cell decomposition $\mathcal{T}^{\max}$ with

$$\mathcal{I}(\Delta) = \{\overline{s} \in M^{|\overline{y}|} : \varphi(\overline{x}; \overline{s}) \text{ crosses } \Delta\},$$

i.e. for every $\psi \in \Psi$ we take

$$\theta_\psi(\overline{y}; \overline{y_1}, \ldots, \overline{y_k}) \equiv (\exists \overline{x}.\psi(\overline{x}; \overline{y_1}, \ldots, \overline{y_k}) \wedge \varphi(\overline{x}; \overline{y})) \wedge (\exists \overline{x}.\psi(\overline{x}; \overline{y_1}, \ldots, \overline{y_k}) \wedge \neg\varphi(\overline{x}; \overline{y})).$$

Moreover, for any finite $S \in M^{|\overline{x}|}$ we have $|\mathcal{T}^{\max}(S)| \leq |\Psi| \cdot |S|^k = O(|S|^k)$.

▶ **Example 16.** In Example 15 we gave an example of the abstract cell decomposition $\mathcal{T}$ weakly definable by $\Psi(x, y_1, y_2)$ for the formula $x = y$ in the model $\mathbb{M}_{\preceq}$. Consider now a set $S = \{v_1, v_3, v_5\}$. In $\mathcal{T}(S)$ there is a set $\{v : v_1 \prec v \prec v_3\} = \{v_2\}$, which is defined by the formula $\psi_2(x; y_1, y_2) \equiv y_1 \preceq x \wedge x \preceq y_2 \wedge x \neq y_1 \wedge x \neq y_2$, where we

plug $y_1 = v_1$ and $y_2 = v_3$. However, if we plug $y_1 = v_1$ and $y_2 = v_5$ we get the set $\Delta = \{v : v_1 \prec v \prec v_4\} = \{v_2, v_3, v_4\} \notin T(S)$.

Consider the formula θ_{ψ_2} given above for defining $\mathcal{T}^{\max}$. In this particular case it says:

$$\theta_{\psi_2}(y, y_1, y_2) \equiv \psi_2(y, y_1, y_2) \wedge (\exists x.x \neq y \wedge \psi_2(x, y_1, y_2)).$$

Observe that we have $\mathbb{M}_{\preceq} \models \theta_{\psi_2}(v_3, v_1, v_5)$, so we have $v_3 \in \mathcal{I}(\Delta)$. In this way we know that the set Δ shouldn't be in $\mathcal{T}(S)$. It is easy to observe, that for the considered abstract cell decomposition $\mathcal{T}$, we have that $\mathcal{T}^{\max}(S) = O(|S|)$ for the induced $\mathcal{T}^{\max}$ definable abstract cell decomposition. ◀

Using the notion of a weakly definable abstract cell decomposition we can define the notion of distality. We define it for single formulas and classes of structures similarly to the notion of NIP. The following definition is an adaptation of [13, Fact 2.9].

▶ **Definition 17.** *A formula $\varphi(\overline{x}; \overline{y})$ with two tuples of parameters $\overline{x}$ and $\overline{y}$ is distal in a class of structures $\mathcal{C}$ if there is a finite set of sentences $\Psi(\overline{x}; \overline{y}_1, \ldots, \overline{y}_k)$ such that for any $\mathbb{M} \in \mathcal{C}$ there is an abstract cell decomposition for $\varphi(\overline{x}; \overline{y})$ weakly definable by Ψ.*

In other words we say that a formula $\varphi(\overline{x}; \overline{y})$ is distal in $\mathcal{C}$ if in every $\mathbb{M} \in \mathcal{C}$ it admits an abstract cell decomposition weakly definable by the same set of formulas. As we observed that every weakly definable abstract cell decomposition $\mathcal{T}$ induces a definable abstract cell decomposition $\mathcal{T}^{\max}$, we can assume that the decomposition in the definition of a distal formula is definable. Moreover, we say that a class of structures $\mathcal{C}$ is distal if every formula $\varphi(\overline{x}; \overline{y})$ is distal in $\mathcal{C}$ and we say that a single structure $\mathbb{M}$ is distal if the class $\{\mathbb{M}\}$ is.

Observe, that similarly to the case of NIP classes of structures, if a formula $\psi(\overline{x}; \overline{y})$ admits a distal cell decomposition weakly definable by a finite set of sentences Ψ in $\mathcal{C}$, then it also admits a distal cell decomposition weakly definable by a finite set of sentences Ψ in the elementary closure $\overline{\mathcal{C}}$.

4.4   Distal combinatorial tools

We start this section with the original statement of Distal cutting lemma.

▶ **Theorem 18** (Distal cutting lemma, [13, Theorem 3.2]). *Let $\mathbb{M}$ be a fixed structure and $\varphi(\overline{x}; \overline{y})$ be a formula admitting a definable distal cell decomposition $\mathcal{T}$ (weakly definable by a finite set of formulas $\Psi(\overline{y}; \overline{y}_1, \ldots, \overline{y}_k)$) with $\mathcal{T}(S) = O(|S|^d)$. Then for any $A \subseteq M^{|y|}$ of size n and any real r satisfying $1 \leq r \leq n$ we can partition $M^{|x|}$ into sets $X_1, \ldots, X_l$ with*

$$l \leq Cr^d,$$

and satisfying the following: for every X_i there are at most $\frac{n}{r}$ tuples $\overline{a} \in A$ such that $\varphi(M; \overline{a})$ crosses X_i. The constant C depends only on the model $\mathbb{M}$ and the formula φ. Moreover, each of the X_i's is an intersection of at most two $\Psi(x; A^k)$-definable sets.

To better understand the statement of Theorem 18 consider it in the case when $\mathbb{M}$ is a graph (possibly ordered and infinite) and $\varphi(x; y) \equiv E(x, y)$. Then Theorem 18 says, that for any set of vertices A we can partition all the vertices into at most Cr^d sets $X_1, \ldots, X_l$ and in every X_i there are at most $\frac{n}{r}$ vertices $a \in A$ such that a has both a neighbor and a non-neighbor in X_i, so in other words X_i is homogenous towards all but $\frac{n}{r}$ vertices from A.

By using compactness in a standard way, we get a version of Theorem 18 for classes of structures.

▶ **Theorem 19** (Distal cutting lemma for classes of structures). *Let $\mathcal{C}$ be a fixed class of structures and $\varphi(\overline{x};\overline{y})$ be a formula admitting for every $\mathbb{M} \in \mathcal{C}$ a definable distal cell decomposition $\mathcal{T}_\mathbb{M}$ (weakly definable by the same finite set of formulas $\Psi(\overline{y};\overline{y}_1,\ldots,\overline{y}_k)$) with $\mathcal{T}_\mathbb{M}(S) = O(|S|^d)$. Then for any $\mathbb{M} \in \mathcal{C}$, any $A \subseteq M^{|y|}$ of size n, and any real r satisfying $1 \leq r \leq n$ we can partition $M^{|x|}$ into sets $X_1, \ldots, X_l$ with*

$$l \leq Cr^d,$$

such that for every X_i there are at most $\frac{n}{r}$ tuples $\overline{a} \in A$ such that $\varphi(M;\overline{a})$ crosses X_i. The constant C depends only on the class of structures $\mathcal{C}$ and the formula φ. Moreover, each of the X_i's is an intersection of at most two $\Psi(x;A^k)$-definable sets.

Proof. Assume by contradiction that the statement is not true for some real r. Therefore, for every $L \in \mathbb{N}$ there is some $n_L \in \mathbb{N}$ and a structure $\mathbb{M}_L \in \mathcal{C}$ together with $A_L \subseteq M_L^{|y|}$ of cardinality at most n_L such that we cannot choose at most Lr^d sets which are intersection of at most two $\Psi(x;A_L^k)$-definable sets and form a desired partition. Without loss of generality, we can choose constants n_L so that they form a non-decreasing sequence.

Observe, that for every L we can write a sentence ξ_L of first-order logic stipulating:

"there is a set A_L^k of at most n_L tuples of length $|y|$ (formally we quantify existentially n_L times over tuples of size $|y|$) with the following property: for every choice of Lr^d sets which are intersection of at most two $\Psi(\overline{x};A_L^k)$-definable sets, if they cover the whole domain, then

at least one of them is crossed by more than $\frac{|A_L^k|}{r}$ $\varphi(\overline{x};A_L^k)$-definable sets".

Consider the theory T of the class $\mathcal{C}$. Extend it by adding all the sentences ξ_L to obtain a new theory T'. By compactness, T' is consistent. Indeed, any finite subset of T' is consistent, as there is a model in $\mathcal{C}$ which satisfies sentences ξ_L for arbitrary large L's.

As T' is consistent, it has a model $\mathbb{M}'$. As $T \subseteq T'$, the model $\mathbb{M}'$ satisfies the assumptions of Theorem 18. However, as $\xi_L \in T'$ for every $L \in \mathbb{N}$, then $\mathbb{M}'$ does not satisfy its conclusion, which gives us a contradiction. ◀

A similar approach can be applied to another distal combinatorial tool - Distal regularity lemma [15, Theorem 5.8]. However, its full statement is more complicated and uses a number of new definitions. Moreover, it is stated for distal models, but a careful analysis of the proof shows that it can be applied also on the level of distal formulas provided that the model we are working with is NIP. Therefore, we decided to present here a simplified statement of this theorem in the setting of (ordered) graphs.

▶ **Theorem 20** (Simplified version of Distal regularity lemma). *Let $\mathcal{C}$ be an NIP class of (ordered) graphs and assume that the formula $\varphi(x;y) \equiv E(x,y)$ is distal in $\mathcal{C}$. Then, there is a constant $c = c(\mathcal{C})$ such that for any $\varepsilon > 0$ and for any graph $G \in \mathcal{C}$ there is a partition $P_1, \ldots, P_l$ of vertices of G for some $l \leq c(\frac{1}{\varepsilon})^c$, such that*

$$\sum |P_i||P_j| \leq \varepsilon |V(G)|^2$$

where the sum is over all $i,j \in [|V(G)|]$ such that P_i and P_j are not homogenous.

A sketch of the proof of Theorem 20 is given in Appendix B. As we mentioned, this is just repeating parts of the argument from [15], but in the simplified setting of graphs.

1	0	1	1
1	0	0	0
1	1	1	0
0	0	0	1
1	0	1	1
1	0	0	1

Figure 4 In this example we have $\mathrm{Prof}(2) = \{1, 4, 5\}$.

5 Abstract cell decomposition for the edge relation in graphs of bounded twin-width

Classes of ordered graphs of bounded twin-width are known to be distal. The proof of this fact is a combination of a few involved results from model theory. Without going into details, if $\widehat{\mathcal{C}}$ is a class of totally ordered graphs of bounded twin-width, then by [7, Theorem 1 and Theorem 3] it is monadically NIP (this is a property of models much stronger than being NIP). Then by [10, Theorem 1.1] we get that since $\widehat{\mathcal{C}}$ is monadically NIP then it is dp-minimal (this is another property of models which we don't define in this paper), and finally by [32, Lemma 2.10] we get that as $\widehat{\mathcal{C}}$ is dp-minimal and totally ordered, then it is distal. However, this proof does not give any explicit abstract cell decomposition for the class $\mathcal{C}$, which means that e.g. we don't have any bound on the parameter d in Theorem 19.

The aim of this section and the main result of this paper is a direct combinatorial proof of the result sketched above for the case of the edge relation. Formally, we prove the following theorem.

▶ **Theorem 21.** *Let $\widehat{\mathcal{C}}$ be a class of ordered graphs with the following property: there is a constant $t \in \mathbb{N}$ depending only on $\widehat{\mathcal{C}}$ such that for every $(G, \preceq) \in \widehat{\mathcal{C}}$ the matrix $M_\preceq(G)$ is t-mixed free. Then the formula $\varphi(x; y) \equiv E(x, y)$ admits in $\widehat{\mathcal{C}}$ a distal cell decomposition weakly definable by a finite set $\Psi(x; y_1, \ldots, y_k)$, where $k = \mathcal{O}(t)$.*

5.1 Proof of Theorem 21

We start by introducing a crucial definition for this part.

▶ **Definition 22.** *Let M be an $m \times n$ $0 - 1$ matrix. We define the **corner-profile** of column j (with $1 \leq j < n$) as the subset $\mathrm{Prof}(j) \subseteq [m - 1]$ such that $i \in \mathrm{Prof}(j)$ if and only if $M[i : (i + 1)][j : (j + 1)]$ is a corner. For a subset of columns $J \subseteq [n - 1]$ we define $\mathrm{Prof}(J) = \bigcup_{j \in J} \mathrm{Prof}(j)$. For an ordered graph $(G, \preceq)$ we identify columns of $M_\preceq(G)$ with vertices of G and therefore assume that the domain of Prof is $V(G)$ and it ranges over subsets of $V(G)$.*

An example of a corner profile of a column is presented in Figure 4.

Before we proceed with the proof of Theorem 21, it is insightful to prove a simpler proposition, which shows some intuitions that will be useful later on. Namely, we prove that vertices with a *large* corner-profile can be efficiently defined using a few parameters.

We use additional notation for an ordered graph $(G, \preceq)$. A vertex v' is the successor of vertex v in $\preceq$ (denoted by $s(v)$) if $v \neq v', v \preceq v'$ and for all $u \in V(G)$ we have either $u \preceq v$

or $v' \preceq u$. Two vertices $v, v' \in V(G)$ are adjacent in $\preceq$ if $v' = s(v)$ or $v = s(v')$. For a vertex v we also denote by $S^k(v)$ the set of k consecutive vertices after v together with v, i.e. $S^k(v) = \{s^l(v) : 0 \leq l \leq k\}$.

▶ **Proposition 23.** *Let $(G, \preceq)$ be a t-mixed free ordered graph and let $v \in V(G)$ satisfy $|\operatorname{Prof}(v)| \geq 2t$. There are a constant f_t depending only on t, a sequence of vertices $(a_1, \dots, a_t) \in \operatorname{Prof}(v)^t$, and a first-order formula in the language of ordered graphs $\varphi(x; y_1, \dots, y_t)$ such that the length of φ is bounded by f_t and $\varphi(G; a_1, \dots, a_t) = \{v\}$.*

Proof. As $|\operatorname{Prof}(v)| \geq 2t$ we can pick a subset $P \subseteq \operatorname{Prof}(v)$ of size t such that every two vertices $w, w' \in P$ are not adjacent in $\preceq$. Let us denote $S = \{u \in V(G) : P \subseteq \operatorname{Prof}(u)\}$. Obviously $v \in S$ and we can easily see that $|S| < 2t$. Indeed, if $|S| \geq 2t$, then we could pick $S' \subseteq S$ of size t such that every two vertices in S' are not adjacent in $\preceq$. Then pairs of rows $\{(w, s(w)) : w \in P\}$ and pairs of columns $\{(u, s(u)) : u \in S'\}$ would induce a t-mixed minor of $M_{\preceq}(G)$. That is a contradiction with the assumption that $(G, \preceq)$ is t-mixed free. Therefore, we take φ which just says that v is the l'th vertex in order $\preceq$ such that $P \subseteq \operatorname{Prof}(v)$. This can be done within the constraints in the statement, as we can use vertices in P as $a_1, \dots, a_t$ and we can define the successor of a given vertex using $\preceq$. Moreover, the size of φ depends only on t (as we discussed that l is smaller than $2t$), so it can be bounded by some constant f_t. ◀

The assumption that one vertex has a large corner-profile is quite strong, in the sense that we can easily imagine a matrix which admits a large mixed minor, but its every column has a small corner-profile (e.g. of a constant size). However, Proposition 23 can be leveraged to the case when a set of consecutive vertices has a large corner-profile, which is formalized in Proposition 24.

▶ **Proposition 24.** *Let $(G, \preceq)$ be a t-mixed free ordered graph and let $v \in V(G)$ satisfy $|\operatorname{Prof}(S^l(v))| \geq 2t$ for some integer l. There are a constant f_t depending only on t, a sequence of vertices $(a_1, \dots, a_t) \in \operatorname{Prof}(S^l(v))^t$, a first-order formula in the language of ordered graphs $\varphi(x; y_1, \dots, y_t)$, and a vertex $v' \in S^l(v)$ such that the length of φ is bounded by f_t and $\varphi(G; a_1, \dots, a_t) = \{v'\}$.*

Proof. Similarly as in the proof of Proposition 23 we can pick a subset $P \subseteq \operatorname{Prof}(S^l(v))$ of size t such that every two vertices $w, w' \in P$ are not adjacent in $\preceq$. Let v_0 be the smallest vertex of G in $\preceq$ and $r_0 = \min\{n \in \mathbb{N} : P \subseteq \operatorname{Prof}(S^n(v_0))\}$, i.e. r_0 is the smallest integer such that first $r_0 + 1$ vertices of G in $\preceq$ have P as a subset of their corner-profile. This is obviously well-defined as $P \subseteq S^l(v)$. Now we set $v_1 = s^{r_0+1}(v_0)$ and $r_1 = \min\{n \in \mathbb{N} : P \subseteq \operatorname{Prof}(S^n(v_1))\}$. We define $v_2, r_2, v_3, \dots$ analogously, until some v_j or r_j is not defined (i.e. $s^{r_j-1}(v_{j-1})$ is the maximal vertex of G in $\preceq$ or $\{n \in \mathbb{N} : P \subseteq \operatorname{Prof}(S^n(v_j))\}$ is empty). We clearly have $j < 2t$. Indeed, if $j \geq 2t$ then the sets of columns $S^{r_0+1}(v_0), S^{r_2+1}(v_2), \dots, S^{r_{2t-2}+1}(v_{2t-2})$ are pairwise disjoint. Moreover, together with pairs of rows $\{(u, s(u)) : u \in P\}$ they induce a t-mixed minor of $M_{\preceq}(G)$, which is a contradiction.

If v_{j-1}, r_{j-1} is the last pair, which was well-defined above, and we also managed to define v_j but not r_j, then we can't have $v_j \preceq v$. That is because we have $P \subseteq \operatorname{Prof}(S^l(v))$, so in this case r_j would be well-defined. Therefore, there is some p such that $v \in S^{r_p}(v_p)$. We also have $s^{r_p}(v_p) \in S^l(v)$. Indeed, we can't have $s^{r_p}(v_p) \prec v$ by the definition of p and if we had $s^l(v) \prec s^{r_p}(v_p)$, then we would have $P \subseteq S^l(v) \subseteq S^{r_p-1}(v_p)$, which contradicts the definition of r_p. We pick the vertex $s^{r_p}(v_p)$ as our v' from the statement. That is because it is either the maximal vertex of G (in which case it is obvious that we can define it even by a formula without parameters) or the predecessor of v_{p+1}. However, v_{p+1} is definable
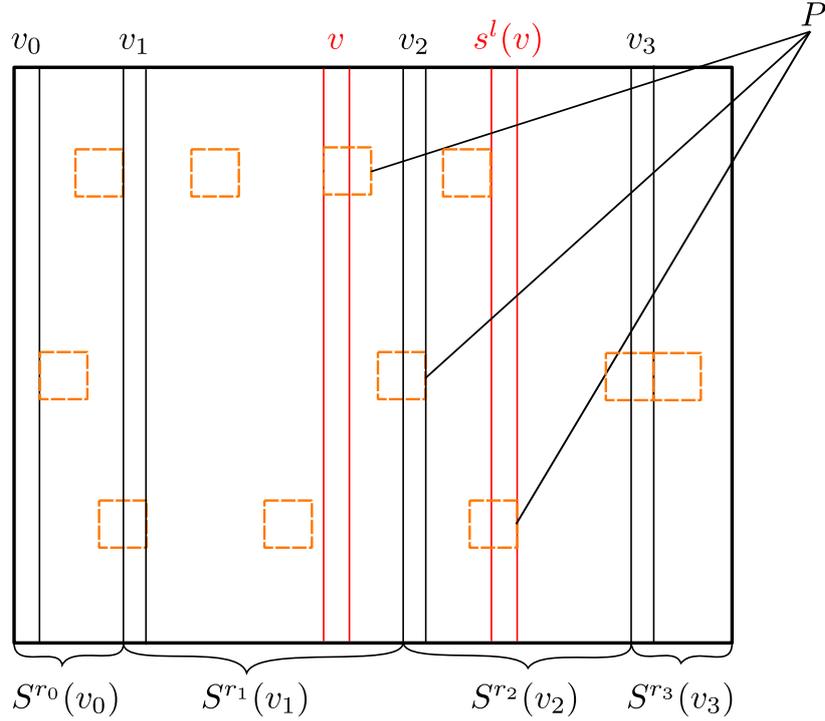
Figure 5 Example of how the v_i are set in the proof of Proposition 24.

by a first-order formula of ordered graphs with $a_1, \ldots, a_t$ being the vertices of P in some order. This is because the definition of each v_i can be easily transformed into a first-order formula of ordered graphs, and the size of formula for v_p can be bounded by some constant f_t because $p \leq 2t$. Finally, as we mentioned in the proof of Proposition 23, the successor relation is definable, so indeed we can define $s^{r_p}(v_p)$ as the predecessor of v_{p+1}. ◀

Let us remark that Proposition 24 assures that if we have a set of consecutive vertices with a large corner-profile, then we can encode *some* vertex in it using a first-order formula. However, we cannot control which vertex from the given set is to be encoded. It turns out not to be that problematic, so we can start the proof of Theorem 21.

Proof of Theorem 21. Let us take a t-mixed free ordered graph $(G, \preceq)$ and some non-empty $A \subseteq V(G)$. Our goal is to present an abstract cell decomposition which is weakly definable by the same set of formulas which does not depend on G nor on A.

Similarly as in the proof of Theorem 4 we can remove from $M_{\preceq}(G)$ rows that correspond to vertices in $V(G) \setminus A$, thus obtaining a t-mixed free $|A| \times |V(G)|$ matrix M. We set v_0 to be the minimal vertex of G in $\preceq$ and $r_0 = \min\{n \in \mathbb{N} : |\mathrm{Prof}(S^n(v_0))| \geq 2t\}$. Next, we set $v_1 = s^{r_0+1}(v_0)$ and $r_1 = \min\{n \in \mathbb{N} : |\mathrm{Prof}(S^n(v_1))| \geq 2t\}$. We proceed this way with $v_2, r_2, v_3, \ldots$. If at some point we manage to define v_l but the set $\{n \in \mathbb{N} : |\mathrm{Prof}(S^n(v_l))| \geq 2t\}$ is empty, then we define $r_l = \max\{n \in \mathbb{N} : s^n(v_l) \in V(G)\}$, i.e. $S^{r_l}(v_l) = \{u \in V(G) : v_l \preceq u\}$. In this way we define $l + 1$ pairs $v_0, r_0, \ldots, v_l, r_l$. This process is presented in Figure 5.

We have that in every $S^{r_j}(v_j)$ there is some v'_j which is first-order definable. This is a bit more tricky than in the proof of Proposition 24, as now any two consecutive rows of M don't necessarily correspond to vertices of $V(G)$ which are adjacent in $\preceq$. However, we are still able to define a corner by explicitly naming both its rows. Therefore, the formula φ_j

which defines v'_j needs to use $2t$ parameters which correspond to t vertices from $\mathrm{Prof}(S^{r_j}(v_j))$ and their successors in $\preceq_{\restriction A \times A}$. There is also a special case for $S^{r_l}(v_l)$, as we can't assume $|\mathrm{Prof}(S^{r_l}(v_l))| \geq 2t$. Nevertheless, the vertex $s^{r_l}(v_l)$ is definable as the maximal vertex of G, so we can set $v'_l = s^{r_l}(v_l)$. As M is t-mixed free, then the size of each φ_j is bounded by some function of t only.

If for any $s^{r_j}(v_j)$ we have $\mathrm{Prof}(s^{r_j}(v_j)) \geq 2t$, then we can define $s^{r_j}(v_j)$ and v_{j+1} (if it exists) as well. Again, we need to use for it formulas with $2t$ parameters of size bounded by some function of t.

By now we defined a few singleton sets with a bounded number of formulas such that any of them uses at most $2t$ parameters from A. Now we will show how to deal with the remaining vertices. Every such vertex u satisfies $w \prec u \prec w'$ for some w, w' already defined in previous steps. Let us denote $I_{ww'} = \{u : w \prec u \prec w'\}$. From the construction we have $|\mathrm{Prof}(I_{ww'})| < 4t$. Moreover, by Corollary 8 we know that there is a subset $A_{ww'} \subseteq A$ of size at most $|\mathrm{Prof}(I_{ww'})| + 1 \leq 4t$ such that whenever two vertices $u, u' \in I_{ww'}$ have exactly the same neighborhood on the set $A_{ww'}$, then their columns in M are equal, so they have the same neighborhood in the whole A. Therefore, we can define sets of vertices between w and w' in $\preceq$, that have specific values on $A_{ww'}$ using formulas with $2t + 2t + 4t$ parameters and of size bounded by some function of t. Finally, we obtain the statement of Theorem 21 with $k = 8t$. The set of formulas Ψ that we use is finite, because every $\psi \in \Psi$ has its size bounded by a function of t only. ◄

5.2 Distal tools for graphs of bounded twin-width

We already know that for every graph G from a class $\mathcal{C}$ of unordered graphs of twin-width bounded by t we can find a total order $\preceq$ on its vertices, such that $(G, \preceq)$ is $2t + 2$-mixed free. Combining this fact with Theorem 21 and Theorem 19 we obtain cutting lemma for graphs of bounded twin-width (without an order).

▶ **Theorem 25** (Cutting lemma for graphs of bounded twin-width). *Let G be a graph of twin-width at most t. There is a constant C depending only on t such that for any $A \subseteq V(G)$ of size n and any real r satisfying $1 \leq r \leq n$ we can partition $V(G)$ into sets $X_1, \ldots, X_l$ with*

$$l \leq Cr^{16t+16}$$

such that for every X_i all but at most $\frac{n}{r}$ vertices $a \in A$ are homogenous towards X_i.

Similarly, using Theorem 20 we can formulate regularity lemma for graphs of bounded twin-width. In this case, we cannot directly apply the bounds on the size of abstract cell decompositions from Theorem 21. The reason is that the bounds in Theorem 20 depend also on VC-dimension of set systems definable by formulas used in a definition of an abstract cell decomposition.

▶ **Theorem 26** (Regularity lemma for graphs of bounded twin-width). *Let G be a graph of twin-width at most t. Then, there is a constant $c = c(t)$ such that for any $\varepsilon > 0$ there is a partition $P_1, \ldots, P_l$ of $V(G)$ for some $l \leq c(\frac{1}{\varepsilon})^c$, such that*

$$\sum |P_i||P_j| \leq \varepsilon |V(G)|^2$$

where the sum is over all $i, j \in [|V(G)|]$ such that P_i and P_j are not homogenous.

6   Conclusions

One of the most important conjectures in structural and algorithmic graph theory is that model checking is FPT on a class of graphs if and only if the class is monadically NIP [1]. As we mentioned, this notion comes from model theory and therefore it is believed that progressing towards this conjecture requires a better understanding of the interplay of model theory and structural graph theory. We believe that studying the notion of distality is a step in this direction.

We can observe the mentioned interplay when we can prove a theorem in two completely different ways – one using tools from model theory, and the other being a direct, combinatorial proof. An example of such theorem is shown in [18]. We are grateful to the Anonymous Reviewer who pointed out that Theorem 26 alternatively can be proven in a purely combinatorial way using [5, Lemma 20]. We believe that studying connections between these two proofs is an interesting goal.

As we mentioned in Section 5 a class of ordered graphs is distal if and only if it has bounded twin-width. However, the distal regularity lemma and the distal cutting lemma does not only apply to graphs of bounded twin-width. As an example, we can consider classes of graphs of bounded expansion equipped with an arbitrary order. They are not distal, but every first-order formula which doesn't use order is distal in them [14]. This leads to a possible generalization of classes of graphs of bounded expansion and bounded twin-width and such a generalization is highly anticipated and sought. In light of these facts, Theorem 25 and Theorem 26 may be seen as a step towards finding common combinatorial properties of the two notions.

──── **References** ────

1   Algorithms, Logic and Structure Workshop in Warwick – Open Problem Session. `https://warwick.ac.uk/fac/sci/maths/people/staff/daniel_kral/alglogstr/openproblems.pdf`, 2016. [Online; accessed 23-Apr-2023].

2   Matthias Aschenbrenner, Alf Dolich, Deirdre Haskell, Dugald Macpherson, and Sergei Starchenko. Vapnik-Chervonenkis density in some theories without the independence property, I. *Transactions of the American Mathematical Society*, 368:5889–5949, 2016. `doi:https://doi.org/10.1090/tran/6659`.

3   Patrick Assouad. Densité et dimension. *Annales de l'Institut Fourier*, 33:233–282, 1983.

4   Pierre Bergé, Édouard Bonnet, and Hugues Déprés. Deciding Twin-Width at Most 4 Is NP-Complete. In Mikołaj Bojańczyk, Emanuela Merelli, and David P. Woodruff, editors, *49th International Colloquium on Automata, Languages, and Programming (ICALP 2022)*, volume 229 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:20, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. URL: `https://drops.dagstuhl.de/opus/volltexte/2022/16359`, `doi:10.4230/LIPIcs.ICALP.2022.18`.

5   Édouard Bonnet, Colin Geniet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. Twin-width II: small classes. In Dániel Marx, editor, *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms, SODA 2021, Virtual Conference, January 10 - 13, 2021*, pages 1977–1996. SIAM, 2021. `doi:10.1137/1.9781611976465.118`.

6   Édouard Bonnet, Colin Geniet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. Twin-width III: Max Independent Set, Min Dominating Set, and Coloring. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 35:1–35:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. URL: `https://drops.dagstuhl.de/opus/volltexte/2021/14104`, `doi:10.4230/LIPIcs.ICALP.2021.35`.

7 Édouard Bonnet, Ugo Giocanti, Patrice Ossona de Mendez, Pierre Simon, Stéphan Thomassé, and Szymon Toruńczyk. Twin-Width IV: Ordered Graphs and Matrices. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 924–937, New York, NY, USA, 2022. Association for Computing Machinery. `doi:10.1145/3519935.3520037`.

8 Édouard Bonnet, Eun Jung Kim, Amadeus Reinald, Stéphan Thomassé, and Rémi Watrigant. Twin-Width and Polynomial Kernels. In Petr A. Golovach and Meirav Zehavi, editors, *16th International Symposium on Parameterized and Exact Computation (IPEC 2021)*, volume 214 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 10:1–10:16, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. URL: `https://drops.dagstuhl.de/opus/volltexte/2021/15393`, `doi:10.4230/LIPIcs.IPEC.2021.10`.

9 Édouard Bonnet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. Twin-width I: tractable FO model checking. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 601–612. IEEE, 2020. `doi:10.1109/FOCS46700.2020.00062`.

10 Samuel Braunfeld and Michael C Laskowski. Characterizations of monadic NIP. *Transactions of the American Mathematical Society, Series B*, 8(30):948– 970, 2021.

11 Bernard Chazelle. *Cuttings*, chapter chapter26. CRC Press, 2018. URL: `https://www.routledgehandbooks.com/doi/10.1201/9781315119335-26`, `doi:10.1201/9781315119335-26`.

12 Bernard Chazelle, Herbert Edelsbrunner, Leonidas J. Guibas, and Micha Sharir. A singly exponential stratification scheme for real semi-algebraic varieties and its applications. *Theor. Comput. Sci.*, 84(1):77–105, Jul 1991. `doi:10.1016/0304-3975(91)90261-Y`.

13 Artem Chernikov, David Galvin, and Sergei Starchenko. Cutting lemma and Zarankiewicz's problem in distal structures. *Selecta Mathematica*, 26(2):25, 2020. `doi:10.1007/s00029-020-0551-2`.

14 Artem Chernikov, Pierre Simon, and Szymon Toruńczyk. Private communication.

15 Artem Chernikov and Sergei Starchenko. Regularity lemma for distal structures. *Journal of the European Mathematical Society*, 20(10):2437–2466, 2018. `doi:10.4171/JEMS/816`.

16 Josef Cibulka and Jan Kynčl. Better Upper Bounds on the Füredi-Hajnal Limits of Permutations. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '17, page 2280–2293, USA, 2017. Society for Industrial and Applied Mathematics.

17 Édouard Bonnet. Twin-width. `https://www.youtube.com/watch?v=P0_2ya3et24`, 2020. [Online; accessed 23-Apr-2023].

18 Jakub Gajarský, Nikolas Mählmann, Rose McCarty, Pierre Ohlmann, Michał Pilipczuk, Wojciech Przybyszewski, Sebastian Siebertz, Marek Sokołowski, and Szymon Toruńczyk. Flipper games for monadically stable graph classes, 2023. `arXiv:2301.13735`.

19 David Haussler. Sphere packing numbers for subsets of the boolean n-cube with bounded Vapnik-Chervonenkis dimension. *Journal of Combinatorial Theory, Series A*, 69(2):217–232, 1995. URL: `https://www.sciencedirect.com/science/article/pii/0097316595900527`, `doi:https://doi.org/10.1016/0097-3165(95)90052-7`.

20 Janos Komlos and Miklós Simonovits. Szemerédi's regularity lemma and its applications in graph theory. *Combinatorica*, 2, 01 1996.

21 Michael C. Laskowski. Vapnik-Chervonenkis Classes of Definable Sets. *Journal of the London Mathematical Society*, s2-45(2):377–384, 04 1992. `doi:10.1112/jlms/s2-45.2.377`.

22 László Lovász and Balázs Szegedy. Regularity partitions and the topology of graphons. In Imre Bárány, József Solymosi, and Gábor Sági, editors, *An Irregular Mind: Szemerédi is 70*, pages 415–446, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. `doi:10.1007/978-3-642-14444-8_12`.

23 Adam Marcus and Gábor Tardos. Excluded permutation matrices and the Stanley–Wilf conjecture. *Journal of Combinatorial Theory, Series A*, 107(1):153–160, 2004. URL: `https://www.`

sciencedirect.com/science/article/pii/S0097316504000512, `doi:https://doi.org/10.1016/j.jcta.2004.04.002`.

24 Jiri Matousek. *Lectures on Discrete Geometry*. Springer-Verlag, Berlin, Heidelberg, 2002.

25 Jiří Matoušek, Emo Welzl, and Lorenz Wernisch. Discrepancy and approximations for bounded VC-dimension. *Combinatorica*, 13(4):455–466, 1993. `doi:https://doi.org/10.1007/BF01303517`.

26 Adam Paszke and Michał Pilipczuk. VC Density of Set Systems Definable in Tree-Like Graphs. In Javier Esparza and Daniel Král', editors, *45th International Symposium on Mathematical Foundations of Computer Science (MFCS 2020)*, volume 170 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 78:1–78:13, Dagstuhl, Germany, 2020. Schloss Dagstuhl–Leibniz-Zentrum für Informatik. URL: `https://drops.dagstuhl.de/opus/volltexte/2020/12747`, `doi:10.4230/LIPIcs.MFCS.2020.78`.

27 Michał Pilipczuk. Introduction to twin-width. `https://www.youtube.com/watch?v=BddeeHWRX_g`, 2021. [Online; accessed 23-Apr-2023].

28 Felix Reidl, Fernando Sánchez Villaamil, and Konstantinos Stavropoulos. Characterising bounded expansion by neighbourhood complexity. *European Journal of Combinatorics*, 75:152–168, 2019. URL: `https://www.sciencedirect.com/science/article/pii/S0195669818301355`, `doi:https://doi.org/10.1016/j.ejc.2018.08.001`.

29 Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972. URL: `https://www.sciencedirect.com/science/article/pii/0097316572900192`, `doi:https://doi.org/10.1016/0097-3165(72)90019-2`.

30 Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1):247 – 261, 1972. `doi:https://doi.org/10.2140/pjm.1972.41.247`.

31 Saharon Shelah. Classification theory and the number of non-isomorphic models, second edition. In *Studies in logic and the foundations of mathematics*, volume 92, Amsterdam, 1990. North-Holland Publishing Co.

32 Pierre Simon. Distal and non-distal NIP theories. *Annals of Pure and Applied Logic*, 164(3):294–318, 2013. URL: `https://www.sciencedirect.com/science/article/pii/S0168007212001790`, `doi:https://doi.org/10.1016/j.apal.2012.10.015`.

33 Terence Tao. Expanding polynomials over finite fields of large characteristic, and a regularity lemma for definable sets. *Contributions to Discrete Mathematics*, 10(1), 2015. `doi:10.11575/cdm.v10i1.62182`.

34 Vladimir Naumovich Vapnik and Alexey Yakovlevich Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16:264–280, 1971.

A   Proof of Lemma 12

In Section 2 for a given graph G and a total ordering σ of $V(G)$ we denoted by $M_\sigma(G)$ the adjacency matrix of G in the order σ. This definition can be lifted to adjacency matrices of arbitrary binary structures, in the following way, which was presented in [9].

Let H be a binary structure over the signature consisting of k binary relations $R_1, \ldots, R_k$ and let σ be a total ordering of the domain of H, say $v_1, \ldots, v_n$. For vertices v_i, v_j and a relation R_l we define

$$
e_{ij}^l = \begin{cases}
0 & \text{if } \neg R_l(v_i, v_j) \wedge \neg R_l(v_j, v_i) \\
1 & \text{if } R_l(v_i, v_j) \wedge \neg R_l(v_j, v_i) \\
-1 & \text{if } \neg R_l(v_i, v_j) \wedge R_l(v_j, v_i) \\
2 & \text{if } R_l(v_i, v_j) \wedge R_l(v_j, v_i),
\end{cases}
$$

and $e_{ij} = (e_{ij}^1, \ldots, e_{ij}^k)$. Finally, we define the adjacency matrix of H in the order σ as the matrix $M_\sigma(H)$ satisfying $M_\sigma(H)[i][j] = e_{ij}$.

Not only the definition of M_σ can be lifted to the case of binary structures, but also the definition of twin-width. We don't present the lifted definition here. Instead, we just remark that by [9, Theorem 14] if for a given binary structure H there is a total order σ on its domain such that $M_\sigma(H)$ is t-mixed free, then the twin-width of H is $2^{2^{O(t)}}$. Therefore, we can show that a binary structure has bounded twin-width, by finding an ordering in which its adjacency matrix is t-mixed free for some t. This happens to be the case for the ordered graphs that we consider in the statement of Lemma 12.

▶ **Lemma 27.** *Let $(G, \preceq)$ be an ordered graph such that the matrix $M_\preceq(G)$ is t-mixed free. Then $M_\preceq((G, \preceq))$ is $2t$-mixed free.*

Before starting the proof, let us stress the difference between $M_\preceq(G)$ and $M_\preceq((G, \preceq))$. The first one is the adjacency matrix of the graph G in the order $\preceq$. In particular, it is a $0-1$ matrix. On the other hand, $M_\preceq((G, \preceq))$ is an adjacency matrix of a binary structure over the signature $\{E, \preceq\}$, so its elements are from the set $(0, -1, 1, 2)^2$. However, as the order $\preceq$ is both a part of the structure and the order which is used for creating $M_\preceq((G, \preceq))$, it is easier to reason about properties of this matrix.

Proof of Lemma 27. Let us assume by contradiction, that $M_\preceq((G, \preceq))$ admits a $(2t, 2t)$-division $(\mathcal{R}, \mathcal{C}) = (\{R_1, \ldots, R_{2t}\}, \{C_1, \ldots, C_{2t}\})$, which is a $2t$-mixed minor. Denote by R_{ij} (respectively C_{ij}) the sum $R_{ij} = \cup_{k=i}^j R_k$ (respectively $C_{ij} = \cup_{k=i}^j C_k$). It is easy to see that on one of the zones $R_{1t} \cap C_{t+1,2t}$ or $R_{t+1,2t} \cap C_{1t}$ the relation $\preceq$ is constant. Therefore, either $(\{R_1, \ldots, R_t\}, \{C_{t+1}, \ldots, C_{2t}\})$ or $(\{R_{t+1}, \ldots, R_{2t}\}, \{C_1, \ldots, C_t\})$ induce a t-mixed minor of $M_\preceq(G)$, which is a contradiction. ◀

The notion of NIP classes of structures can be extended to much stronger notion of monadically NIP classes. Roughly speaking a class $\mathcal{C}$ is monadically NIP if whenever we add a number of unary predicates to every structure in $\mathcal{C}$ thus obtaining a class $\overline{\mathcal{C}}$, then $\overline{\mathcal{C}}$ is NIP. We don't define the notion of monadically NIP classes of structures formally, because for the proof of Lemma 12 it is enough to know that this is a more restrictive notion than NIP. Knowing that we can proceed to the proof of Lemma 12.

Proof of Lemma 12. By Lemma 27 and [9, Theorem 14] we know that the class $\widehat{\mathcal{C}}$ of ordered graphs has bounded twin-width. By [7, Theorem 11] and the equivalence of conditions (1) and (2) of [7, Theorem 3] we get that $\widehat{\mathcal{C}}$ is monadically NIP and hence NIP. ◀

B   Sketch of the proof of Theorem 20

In this section whenever we talk about graphs we assume that they can be ordered. Before we sketch the proof of Theorem 20 we need to define a few auxiliary notions.

Let us assume that we have a class of graphs $\mathcal{C}$ such that the formula $\varphi(x; y) \equiv E(x, y)$ admits an abstract cell decomposition in $\mathcal{C}$ weakly definable by $\Psi(x; y_1, \ldots, y_l)$. For any $G \in \mathcal{C}$, any $\overline{d} \in G^l$, and any $\psi \in \Psi$ we denote by $C_{G, \psi, \overline{d}}$ the subset of vertices of G defined by $\psi(G; \overline{d})$ and call it a *chamber*. If in addition $\overline{d} \in B^l$ for some $B \subseteq V(G)$ then we say that $C_{G, \psi, \overline{d}}$ is a *B-definable* chamber.

For a chamber C and a set B we denote by $C^{\#}(B)$ the set of all $b \in B$ such that C and $\{b\}$ are not homogenous. Note that $C^{\#}(G)$ is a definable set (by a formula depending just on the formula ψ defining C).

For $B \subseteq V(G)$ we say that a chamber C is *B-complete* if C is B-definable and $C^{\#}(B) = \emptyset$.

Since Ψ weakly defines an abstract cell decomposition of φ in $\mathcal{C}$ then for any $G \in \mathcal{C}$, any finite $B \subseteq V(G)$, and any $a \in V(G)$ there is a B-complete chamber C such that $a \in C$. In particular for any finite $B \subseteq V(G)$ the union of all B-complete chambers covers $V(G)$.

For a positive $r \in \mathbb{R}$ and a $G \in \mathcal{C}$ we say that a family $\mathcal{F}$ of chambers is a *1/r-cutting* if $V(G)$ is covered by $\mathcal{F}$ and for every $C \in \mathcal{F}$ we have $C^{\#}(G) \le |V(G)|/r$.

We also need the following fact, which can be seen as a special case of the VC-theorem.

▶ **Lemma 28** (Simplified version of [15, Fact 2.2]). *For any $k > 0$ and $\varepsilon > 0$ there is $n = \mathcal{O}(k(\frac{1}{\varepsilon})^2 \log \frac{2}{\varepsilon})$ satisfying the following. For any finite set X and a family $\mathcal{F}$ of subsets of X of VC-dimension $\le k$ there are some $x_1, \dots, x_n \in X$ such that for any $S \in \mathcal{F}$ if $|S| \ge \varepsilon |X|$ then $S \cap \{x_1, \dots, x_n\} \ne \emptyset$.*

After explaining these additional notions we can prove Lemma 29, which is an important step in the direction of Theorem 20.

▶ **Lemma 29** (compare with [15, Claim 3.5]). *There is a constant K depending only on $\mathcal{C}$ such that the following holds. For any positive r and for any $G \in \mathcal{C}$ there is a set $S \subseteq V(G)$ of size at most $Kr^2 \log 2r$ such that the family of all S-complete chambers is a $1/r$-cutting.*

Proof. Consider the family of sets

$$\mathcal{D} = \{C^{\#}(G) : C \text{ is a } G\text{-definable chamber}\}.$$

This is a definable family, so it has VC-dimension bounded by a constant that depends only on $\mathcal{C}$. By applying Lemma 28 with $\varepsilon = \frac{1}{r}$ we obtain a subset $S \subseteq G$ of size at most $Kr^2 \log 2r$ such that for every G-definable chamber C if $|C^{\#}(G)| \ge |V(G)|/r$ then $S \cap C^{\#}(G) \ne \emptyset$. Since $C^{\#}(S) = \emptyset$ for any S-complete chamber C, we are done. ◀

By following the proof of [15, Theorem 3.6] and using Lemma 29 we can prove the following theorem.

▶ **Theorem 30.** *Let $\mathcal{C}$ be an NIP class of graph such that $\varphi(x; y) \equiv E(x, y)$ is distal in $\mathcal{C}$. Then, there is a constant δ depending only on $\mathcal{C}$ and a pair of formulas $\psi_1(x, \overline{z}_1), \psi_2(y, \overline{z}_2)$ such that for every $G \in \mathcal{C}$ there are $\overline{c}_1 \in V(G)^{|z_1|}, \overline{c}_2 \in V(G)^{|z_2|}$ such that $|\psi_1(G, \overline{c}_1)|, |\psi_2(G, \overline{c}_2)| \ge \delta|V(G)|$ and the pair of sets $\psi_1(G, \overline{c}_1), \psi_2(G, \overline{c}_2)$ is homogenous.*

Now we start the sketch of the proof of Theorem 20. The idea is to consider for any $G \in \mathcal{C}$ rectangular partitions of $V(G)^2$, i.e. partitions where every part is of the form $S_1 \times S_2$ for $S_1, S_2 \subseteq V(G)$. In particular, we consider partitions that are (ψ_1, ψ_2)-definable over some $A \subseteq V(G)$, for some formulas $\psi_1(x, \overline{z}_1), \psi_2(y, \overline{z}_2)$. It means that for every $S_1 \times S_2$ in our partition S_1 (respectively S_2) can be expressed as a finite Boolean combination of sets from the family $\{\psi_1(V(G), \overline{a}_1) : \overline{a}_1 \in A^{|z_1|}\}$ (respectively $\{\psi_2(V(G), \overline{a}_2) : \overline{a}_2 \in A^{|z_2|}\}$). For a rectangular partition $\mathcal{P}$ of $V(G)^2$ we can define its defect as:

$$\text{def}(\mathcal{P}) = \sum_{\substack{S_1 \times S_2 \in \mathcal{P} \\ S_1 \text{ and } S_2 \text{ not homogenous}}} |S_1||S_2|.$$

The idea behind the proof of Theorem 20 is to start with the rectangular partition consisting of just one part $V(G) \times V(G)$ and then iteratively decrease its defect using Theorem 30 (possibly applied to subgraphs of G). In this way we obtain a rectangular partition of $V(G)^2$ with a low defect which is (ψ_1, ψ_2)-definable over some set A of size polynomial in $1/\varepsilon$. This process is depicted in [15, Claim 5.7]. We finish the proof by observing, that the statement of Theorem 20 follows if we take the partition of $V(G)$ into (ψ_1, ψ_2)-types over A, which is formally stated in the proof of [15, Theorem 5.8].